图书情报工作
第63卷 第2期 2019年1月

唐诗知识图谱的构建及其智能知识服务设计*

周莉娜[1] 洪亮[1] 高子阳[2]
[1] 武汉大学信息管理学院 武汉 430072 [2] 武汉大学计算机学院 武汉 430072

摘要：[**目的/意义**] 立足于当前大数据环境下的唐诗知识服务需求，以大规模唐诗数据为基础构建唐诗知识图谱并提供智能知识服务，推动人工智能环境下唐诗知识管理和知识服务方式的创新。[**方法/过程**] 本文在对领域知识服务需求调研的基础上，设计领域知识服务驱动的唐诗本体模型，然后利用从 Web 上爬取的多源异构数据，采用知识抽取、知识融合、知识推理等技术自动构建唐诗知识图谱，统一表示和组织唐诗领域数据，实现对大规模唐诗数据的语义化处理。[**结果/结论**] 本文设计基于唐诗知识图谱的智能知识服务平台 KnowPoetry，提供唐诗领域的知识探索、时空轨迹、语义查询等智能化知识服务，推动人工智能环境下唐诗数字人文研究方法的创新转型。

关键词：唐诗知识图谱 智能知识服务 数字人文 知识建模
分类号：G203 I222.7
DOI：10.13266/j.issn.0252-3116.2019.02.003

伴随着因特网的发展，数字人文领域普遍利用数字图书馆、数字中心、万维网联盟（W3C）、GIS 等技术建立了大量的电子库资源，作为数字人文研究者关注的重要领域之一，唐诗领域也积累了大量的数据资源。唐诗不仅代表了中华传统文化的发展高峰，同时也对世界上许多民族和国家的文化发展产生了很大影响。在当前一带一路文化互通共荣的历史背景下，唐诗数据更是全球关联开放数据云（Linked Open Data Cloud）中保障跨语言资源有效获取的重要数据来源[1]。

目前互联网上存在着大量的结构化及非结构化唐诗知识库、文献库等，组织结构松散，知识体系复杂，资源异构分布[2]；而且唐诗在用词、句式、语法、对仗等方面要求较高，有着特殊的文法和习惯，传统的文本挖掘技术难以处理这样的古文文本，因此给大数据环境下的唐诗知识组织和利用带来了极大的挑战。此外，人工智能时代唐诗领域对于基于知识图谱的智能知识服务需求愈加迫切，而现有的唐诗智能应用对唐诗内容的理解不够深入，缺乏对唐诗领域知识进行系统化的表示和建模。以 RDF 三元组为基本表示框架的知识图谱技术描述了知识及知识之间的关联，可以对多源异构的唐诗知识进行统一建模、组织和利用，实现唐诗领域知识利用的四大创新：①厘清基于知识关联的唐诗领域知识服务需求，解决领域实际问题，推动唐诗研究朝着语义化、智能化的方向发展；②探索领域知识服务驱动的唐诗本体建模方法，实现概念体系融合，促进本体更好地支持领域知识服务；③支撑大规模唐诗知识图谱的自动构建，使得基于唐诗知识图谱的智能化应用成为可能，促使传统文化融入"一带一路"公共文化服务建设；④提供基于唐诗知识图谱的智能知识服务应用策略，实现由信息资源向知识化资源的转换，为中华传统文化带来新的发展空间。

因此，本文结合对唐诗领域知识服务需求的调研，首先设计领域知识服务驱动的唐诗本体模型；然后利用唐诗知识图谱来实现对唐诗领域海量数据的语义化处理和存储，搭建面向领域知识服务的唐诗智能服务平台 KnowPoetry，从而支持唐诗领域的知识探索、时空轨迹、语义查询、知识推理等智能化知识服务，推动人工智能环境下唐诗知识组织和知识利用方式的创新，探索唐诗领域数字人文研究范式的进一步转变和升级。

* 本文系国家重点研发计划"科学大数据管理系统"子课题"图数据管理关键技术及系统"（项目编号：2016YFB1000603）和教育部人文社会科学重点研究基地重大项目"大数据资源的语义表示与组织研究——面向文化遗产领域"（项目编号：16JJD870002）研究成果之一。

作者简介：周莉娜（ORCID：0000-0003-3386-3892），硕士研究生；洪亮（ORCID：0000-0002-1466-9843），保密管理系副主任，副教授，博士，通讯作者，E-mail：hong@whu.deu.cn；高子阳（ORCID：0000-0001-7300-3200），本科生。

收稿日期：2018-07-16 修回日期：2018-11-13 本文起止页码：24-33 本文责任编辑：王传清

周莉娜，洪亮，高子阳. 唐诗知识图谱的构建及其智能知识服务设计[J]. 图书情报工作,2019,63(2):24-33.

1 国内外相关研究

1.1 唐诗数字人文研究进展

与唐诗领域相关的数字人文研究方兴未艾,除了对诗歌本身的研究,还包括与此相关的历史、地理、艺术等领域,其发展历程主要经过了由数字化的电子文本库到结构化的主题数据库,再到语义化的关联知识库三个阶段。

早期的数字人文项目主要致力于数据库建设与数字化,出现了如全唐诗库、中华诗词库等电子文献库,如哈佛大学的"中文哲学电子书计划"CText(Chinese Text Project)[3]着眼于中文语言文献的数字化,收录了众多唐宋时期经典著作的电子版和中文善本的影印资料。因此,这一时期的电子文献库主要是非结构化的电子文本数据和图片数据,这些数据库提供浏览和简单的关键词匹配检索等,尚未实现语义化的知识服务和智能检索等功能。唐诗相关的研究依旧遵循通过收集、遴选和比对来构建少量诗歌知识间逻辑关系的研究方法,难以突破传统定性研究方法的局限。

随着结构化的关系数据库技术的发展,数字人文领域出现了诸多结构化的主题数据库,典型代表是由哈佛大学郝若贝(Robert M. Hartwell)与北京大学等机构合作创建的"中国历代人物传记数据库"(China Biographical Database Project,CBDB)[4],包含了大量唐代诗人或文学家群体的传记信息,其结构化的方式有助于学者进行大规模分析。此外,哈佛大学与复旦大学合作创建了"中国历史地理信息系统"(China Historical Geographic Information System,CHGIS)[5],该数据库通过建立连续的时间序列,描述地名、行政建制和其他基础地理要素随时间的变化情况,提供检索和查询中国基础历史地理信息。在国内,王兆鹏搭建了唐宋文学编年地图"搜韵"[6]平台,将诗人的时空轨迹分布信息可视化,并提供关键词匹配的检索入口。与此同时,主题数据库的建立为学者群体提供了研究平台和研究工具的支持,出现了诸多运用数字工具理解中国历史和文化的研究成果。如由莱顿大学的魏希德(Hilde De Weerdt)等创建的古籍半自动标示平台MARKUS[7],支持将中文古文文本自动加上任务、地点以及使用者定制概念等标签的功能;耶鲁大学的"广厦千万间"项目(Ten Thousand Rooms Project)[8]则利用互联网的社区属性,为学者们提供一个就公版文献展开协作研究的平台,比如为某部唐诗集创建一个注释版本;J. W. Chen[9]开展了唐诗与世说新语的文本挖掘研究;P. Jason[10]进行了对宋朝佛教诗人的GIS/网络分析等。

当前,随着关联语义技术和人工智能技术的发展,出现了语义化的关联知识库,如全球开放关联数据库(LOD)、维基百科知识图谱(DBpedia)等全球通用的大规模跨语言知识库,其中涉及中国古代文学和历史地理的部分语义化的数据。而图档博机构在开展数字人文研究上具有强大的数据优势[11],因此近几年也出现了不少图档博机构利用馆藏资源开发的一系列关联语义知识库。如上海图书馆已经建立的盛宣怀知识库、家谱知识库、古籍循证平台和基础知识库(历史人物规范库、历史地理知识库、历史纪年知识库、历史事件知识库)等。其中,在构建家谱知识库时,上海图书馆基于书目框架(BIBFRAME),利用本体建模方法设计了家谱本体,用以解释家谱资源的文献特征和内容属性,增强内容之间的语义关联[12];在构建历史地理知识库中,采用知识本体方法设计了历史地理数据时空模型,以满足图书馆数字人文建设项目中的开放应用需求[13]。

1.2 唐诗智能知识服务研究

目前唐诗领域的知识服务应用是数字人文领域和计算机领域的研究者共同关注的热点问题。唐诗智能知识服务的应用主要分为诗词机器人、智能问答和诗词知识图谱三大类。

在智能写诗领域,国外较早地提出并实现了智能写诗软件的应用,但主要是基于外文诗歌创作风格的自动写诗应用。例如,早在1959年,德国卢茨就制作出最早的写诗软件,其后还有人设计出RACTER和PROSE等作诗系统。国外典型的写诗软件有Desktop Poet(桌面诗人)[14],是一款用来支持英文诗歌创作的工具;Google设计的一款LED互动装置Poetrics,利用声音检索技术和语音平台将语音素材转换为诗歌;英国SwiftKey公司创建的自动化诗人"流畅",用以创作类似莎士比亚十四行诗风格的作品。随着人工智能技术不断发展,国内出现了大量诗歌创作系统,如清华大学孙茂松提出了基于深度学习的自动写诗算法,并设计了基于格律的自动写诗机器人"九歌"[15]。此外还有利用模板拼接写诗的在线作诗软件[16]、IBM"偶得"等;基于主题写诗的有百度"为你写诗""微软绝句"[17]和"微软小冰"[18]等。然而机器创作的诗歌含义和意境较差,并不能满足唐诗领域深度的知识服务需求。

随着语义关联技术的深入发展，通用的搜索引擎开始提供诗歌知识的智能问答功能，即通过对海量诗歌信息的整合加工，根据自然语言对诗歌信息进行询问描述，直接提供答案。如百度搜索“举头望明月”的下一句，搜索引擎便会直接返回答案“低头思故乡”。另外，该领域也出现了譬如“诗歌小强”等专门化的诗歌智能问答系统，向用户提供基于关联知识图谱的语义检索服务。但上述智能问答系统针对复杂的自然语言查询，仍无法摆脱传统的关键词匹配技术，有时难以直接返回给用户想要的答案。

知识图谱技术的出现为唐诗领域提供了新的知识服务思路，与唐宋诗词文化相关的领域知识图谱构建和知识服务研究也成为学界的主流。如北京师范大学开发的基于知识图谱的全唐诗检索与可视化平台“唐诗别苑”[19]，实现了语义检索和知识图谱可视化，但缺乏对唐诗领域知识阐释性、关联性和历史性的理论模型框架研究，因此难以支持深层次复杂化的唐诗知识关联挖掘和知识推理证伪研究。北京大学信息管理系KVision实验室开发了宋代学术传承知识图谱[20]，从CBDB中抽取宋代人物之间的学术传承关系和部分亲属关系，应用知识图谱实现对数据的展示和查询，提供动态的、可视化的历史知识探索与发现。但目前该图谱所呈现的关系限于学术传承关系和亲属关系，通用学术与社会关系的本体化和语义化尚未实现。

纵观国内外研究，已经涉及唐诗知识图谱、唐诗知识服务等主题。但在面向唐诗知识图谱构建及领域智能知识服务的应用方面存在诸多不足：没有系统地对唐诗领域知识进行表示和建模；没有针对多源异构的大规模知识图谱的自动构建方案；缺乏对唐诗领域智能知识服务需求的深入理解和支持；没有从量化的、客观的、动态的视角实现语义检索、关联分析、知识推理等智能唐诗知识服务平台的设计方案。因此，本文力图突破上述研究难点，以唐诗知识图谱的构建及其领域知识服务为范本，探索人工智能时代数字人文研究方式的转变与创新。

2 领域知识服务驱动的唐诗本体建模

基于语义Web的唐诗领域知识表示模型是构建唐诗知识图谱的基础，是机器可理解唐诗语义知识的前提，同时也是支持唐诗智能服务平台建设的基础。现有的唐诗语义关联知识库普遍缺乏对唐诗领域知识进行系统化的表示和建模，因此本文结合对唐诗领域知识服务需求的调研，设计领域知识服务驱动的唐诗本体模型。

2.1 唐诗领域知识服务需求

唐诗知识图谱的构建必须立足于当前大数据环境下基于学者导向的研究需求与基于知识智能的服务需求，从而指导我们设计和搭建支持智能化的唐诗领域知识服务的平台。当前唐诗研究的理论和方法已然成熟，唐诗知识涉及诗学、文献学和史学三大领域的交叉，不同的研究领域对知识图谱的构建需求既有相关性也有一定区别：①诗学研究在形式上注重格律、用词、句式、语法、对仗等，在内容上侧重唐诗的情感意象属性，然而针对唐诗文本的词典构建还属于空白，因此克服唐诗文本词问题，自动化地建立诗歌意象情感的关联成为了当前诗学研究的一大难题。②文献学研究则关注收录唐诗的历代古籍文献的版本源流、编纂校勘、内容真伪等，然而很多唐诗相关的文献记载已很难找到。唐以后的各朝代对唐诗的收录编纂多有差别，因此，克服文献缺失的问题，建立诗人的诗作风格属性与诗歌意象情感等内容属性的关联，判断诗作真伪，用以支持唐诗文献的考据证研究是文献学研究的一大难点。③史学的研究主要挖掘与唐诗相关的唐代政治、经济、文化、民俗地理知识。其中最主要的便是历史地理数据的映射、转换问题，如古今地名的演变发展、公元纪年与历史纪年的转换等，从中挖掘诗人的生平经历和创作发展轨迹是史学研究需要突破的难点。

因此，我们在充分分析领域知识服务需求的前提下，针对上述研究热点和难点，结合从Web上爬取唐诗领域的多源异构数据特点，设计面向诗学、文献学和史学三大特定领域研究需求的唐诗本体，包括面向诗学和文献学应用的诗歌-诗人建模和面向史学应用的时空经历建模。

2.2 面向诗学和文献学应用的诗歌-诗人本体模型

作为中国传统古典文学的一种，唐诗文本有着特殊的文法和习惯，其中包含意象、意境、情感、典故、主题等内容属性，这些属性之间的关系往往都是非表征的关系。因此，本文提出建立“诗歌”实体模型，用以支持面向诗学的意象关联分析。同时，诗歌和诗人之间是作者关系，为此建立“诗人”实体模型。通过诗歌和诗人信息判断唐诗版本的源流发展、真伪信息等便是文献学研究的重要内容。基于RDF三元组的表示方法是构建知识图谱时常用的方法，它提供了知识建模的灵活性；同时我们需要基于RDF的诗歌实体模型的基础上，扩展诗歌相关的实体、属性及实体关系，用以支持唐诗证伪推理的应用。虽然RDF提供了知识

建模的灵活性，但其扩展性不强，而 RDF Schema (RDFS)则在 RDF 基础上提供了一个术语、概念等的定义方式，以及哪些属性可以应用到哪些对象上。由此，本文基于 RDF/RDFS 的三元组表示方法设计面向诗学和文献学应用的诗歌-诗人本体，如图 1 所示：

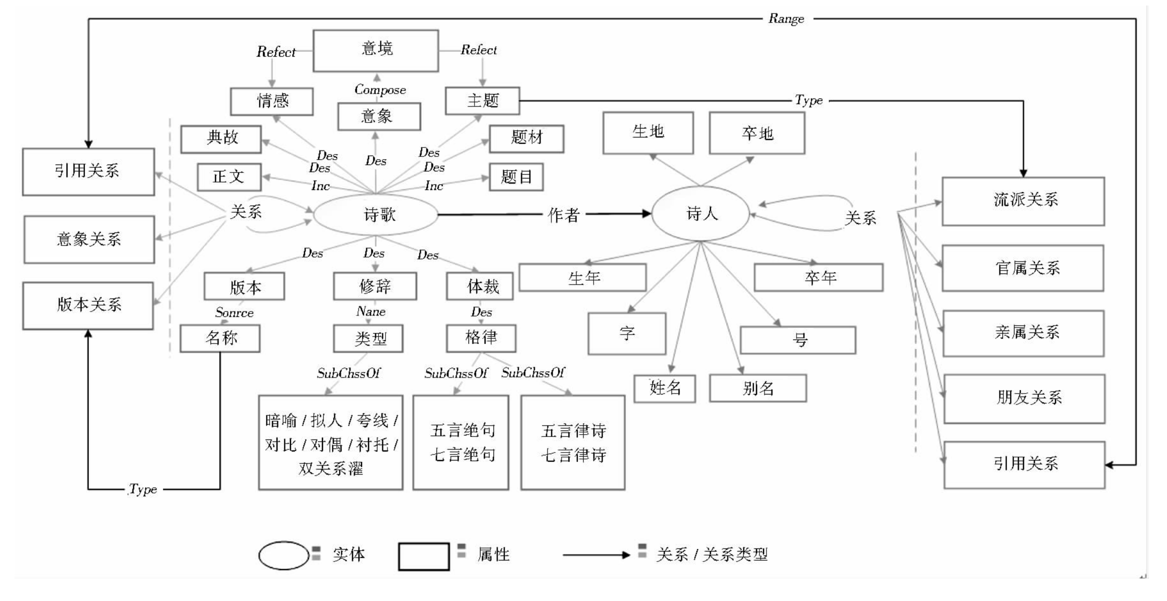


图 1　面向诗学和文献学应用的诗歌-诗人二元本体模型

在这个二元本体模型中，以诗歌为中心实体，诗歌的属性分为 *Inc*：内容性属性和 *Des*：描述性属性，属性之间又存在着上下位类的关联，用 RDFS 的词汇进行约束，如诗歌的体裁属性 *rdfs*：*subClassOf*"七言绝句"是 *rdfs*：*Class*"绝句"的子类。此外，诗歌的诸多意象组成意境，意境则又表现了诗歌的情感属性和主题属性。诗歌-诗歌间的关系则有意象关联关系、版本关联关系和引用关系。在以"诗人"为中心实体的视角下，定义诗人的属性和诗人-诗人的关系，包括引用关系、流派关系、亲属关系、朋友关系、官署关系等。

RDFS 为 RDF 加上了一层约束后，能够实现推理机制，主要表现在诗人的引用关系、流派关系和诗歌的版本关系上。其中，诗人的引用关系抽取则主要通过 Range："诗歌"来体现。首先通过互相提及这一规则来确立确信无疑的互引用，其次定义诗的标题和正文中只要提到过对方，那么两者之间便增加一次引用。同时建立一个对应的引用实体，或称引用事件，这个实体连接两位诗人和对应的诗歌。一首诗如果多次提到对方，只算一次引用。通过这一规则，最终得到诗人间的引用关系，如"刘禹锡" *rdfs*：*author*"酬乐天扬州初逢席上见赠"和"白居易" *rdfs*：字"乐天"，由此得出"刘禹锡" *rdfs*：*rel*"白居易"（引用关系）。而诗歌的版本关系则通过收录编纂诗歌的文献来源确定；同时诗人的流派关系则通过定义和约束诗歌的主题所属的类型建立关联。

《全唐诗》版本中收录《兰溪棹歌》的作者为戴叔伦（732－789 年，字幼公，江苏人），戴诗风格清新流利，以田园山水派诗作居多。这一版本中的《兰溪棹歌》主题属性是歌颂自然，同时表现了隐逸生活的闲适情感属性。而在《明诗三百首》版本中却把《兰溪棹歌》归入了另一个作者汪广洋（？－1380 年，名朝宗，江苏人）的名下，未见收录戴叔伦此诗，由此便出现了疑点。根据构建的诗歌-诗人二元本体模型，我们可以设定推理规则进行版本证伪：同一诗人诸诗作的各项属性基本一致，尤其是诗作风格或者诗作流派较为固定，诗人的流派关系则通过定义和约束诗歌的主题所属的类型建立关联。若某个诗人的诗集中混入了他人作品，可从风格上进行判断。即：

IF（*Poet_School*："田园山水派"）

Then（*Poetry_WroteByPoet_Theme*："田园山水"）

通过诗歌-诗人二元本体模型中诗歌实体的主题、情感属性同诗人实体的流派属性之间的关联，我们可以得出《兰溪棹歌》的《全唐诗》版本是正确的。同时抽取到诗人的时空经历信息，发现汪广洋曾到过兰溪，也曾另作一首《兰溪棹歌》，主题以描绘渔家生活和军旅生活为主。因此，通过上述的规则示例，我们可以发现不同诗集版本的真伪，从而对一首诗同时存在于几个诗人名下的错误辑录情况进行推理和判断，如图 2 所示：

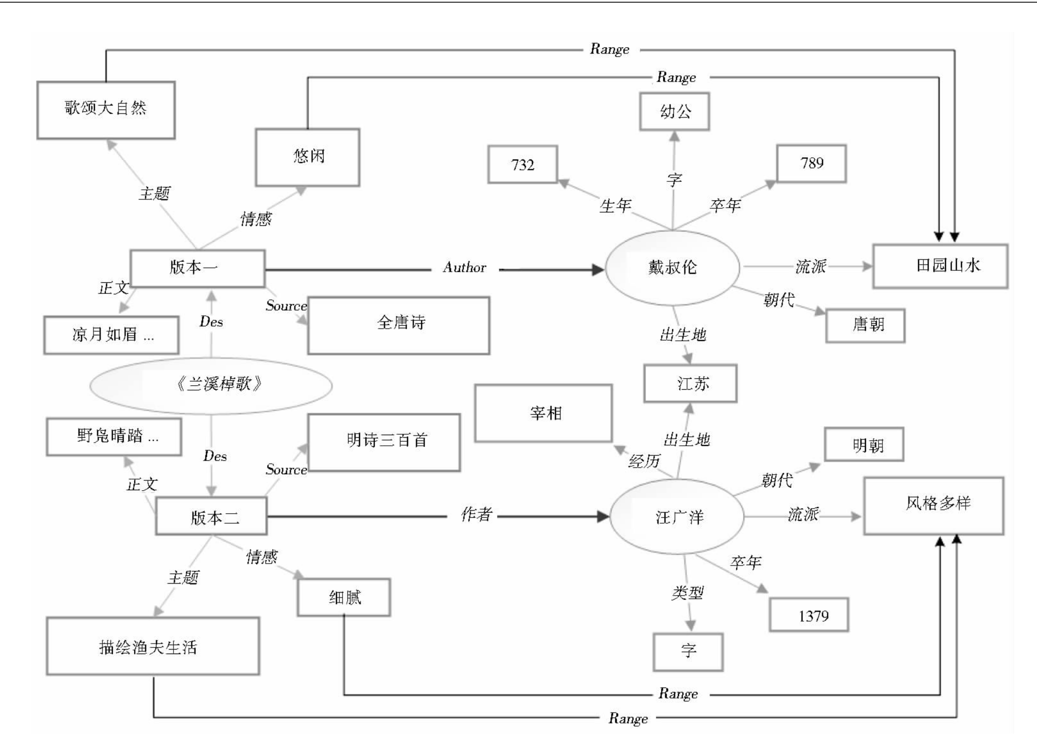


图 2 《兰溪棹歌》版本关系知识推理示例

2.3 面向史学的时空经历本体模型

诗人的生平经历不仅是分析诗人创作发展的重要数据,同时诗人的个人经历也折射出唐代社会的变迁,对研究唐代社会的历史风貌具有重要的意义。因此本文将诗人经历作为第三类实体,其下划分 4 个规范类,即人物、地点、时间和事件来描述一个完整的经历,构建诗人的时空经历本体。首先,地点对应着现实世界中存在着或存在过的空间实体,时间则包括中国历史纪年与公元纪年的对照与转换[21],因此构建地点模型和时间模型来实现数据对象在互联网环境下的访问、定位和关联。其次,事件具有时序化的特点,事件概念之间存在包含、因果、连续、继承等关系,此外,事件的发生也会影响诗人诗歌创作内容的不同。因此,构建诗人经历时空序列模型,来反映诗人经历的时序变化特征,如图 3 所示:

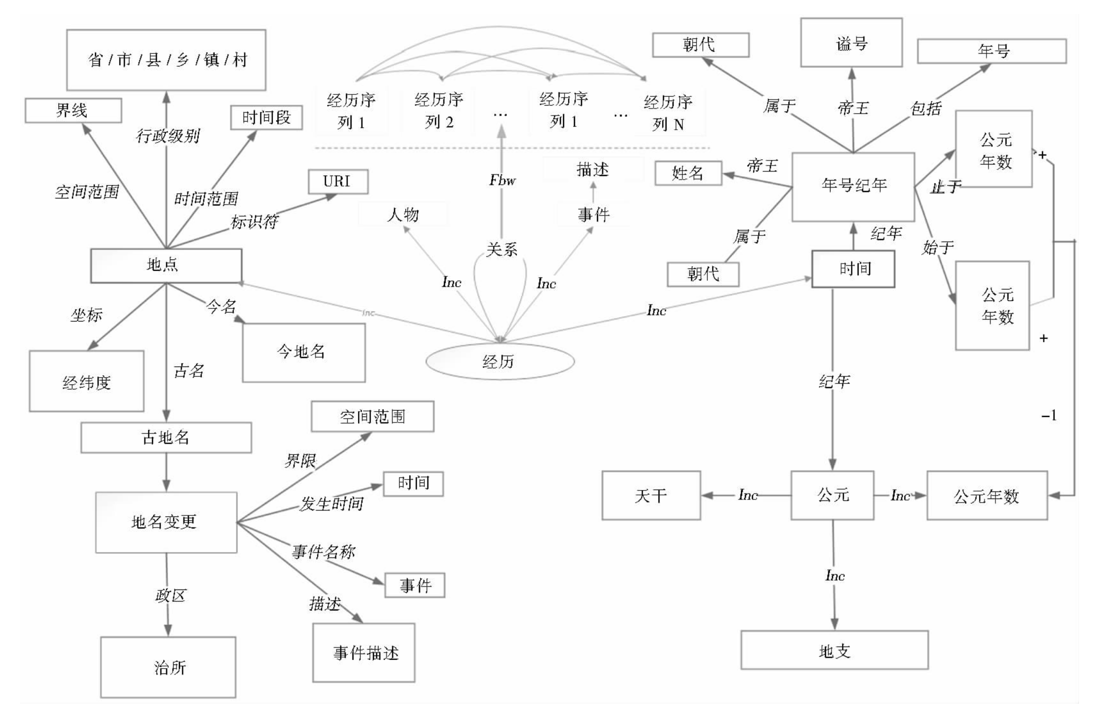


图 3 面向史学的时空经历本体模型

3　唐诗知识图谱构建

结合对领域需求的调研分析，依据第 2 部分领域知识服务驱动的唐诗本体模型，构建唐诗知识图谱如图 4 所示，其主要过程是利用获取的多源唐诗数据，通过知识抽取和知识融合等技术，构建唐诗知识图谱，以 RDF 三元组的形式存储于图数据库中并在此基础上搭建智能化的唐诗知识服务平台。

3.1　数据获取

从 Web 上爬取唐诗领域的多源异构数据，包括百科类网站、中文诗歌网站、人名地名辞典、时空坐标数据等，进行数据转换和加工。

3.2　知识抽取

诗人、诗歌、经历等作为实体都有着各自的属性，同时互相之间存在着各种各样的关系，而这些主要通过知识抽取来补充和丰富，因此唐诗文本的知识抽取主要分为实体的属性抽取和关系抽取。但是目前针对唐诗文本的词典构建还属于空白，为此我们建立信息熵的分词模型来解决唐诗文本的分词问题，实现唐诗文本的自动抽取，如经历抽取、意象情感抽取、引用关系抽取等。

3.3　知识融合

唐诗知识图谱的融合过程就是采用自然语言处理、机器学习等人工智能技术对得到的多源数据进行区分和消歧，建立不同的实体集。根据抽取到的关系类数据，对总数据库中的各部分实体进行链接，将不同的信息类型根据不同模式（属性、虚实体、谓语）嵌入到数据库中。

3.4　图谱管理

通过实体消歧、实体链接等方法进行唐诗知识融合之后，我们将抽取得到的知识库以 RDF 三元组形式进行存储，并使用图数据库系统 gStore 进行管理。

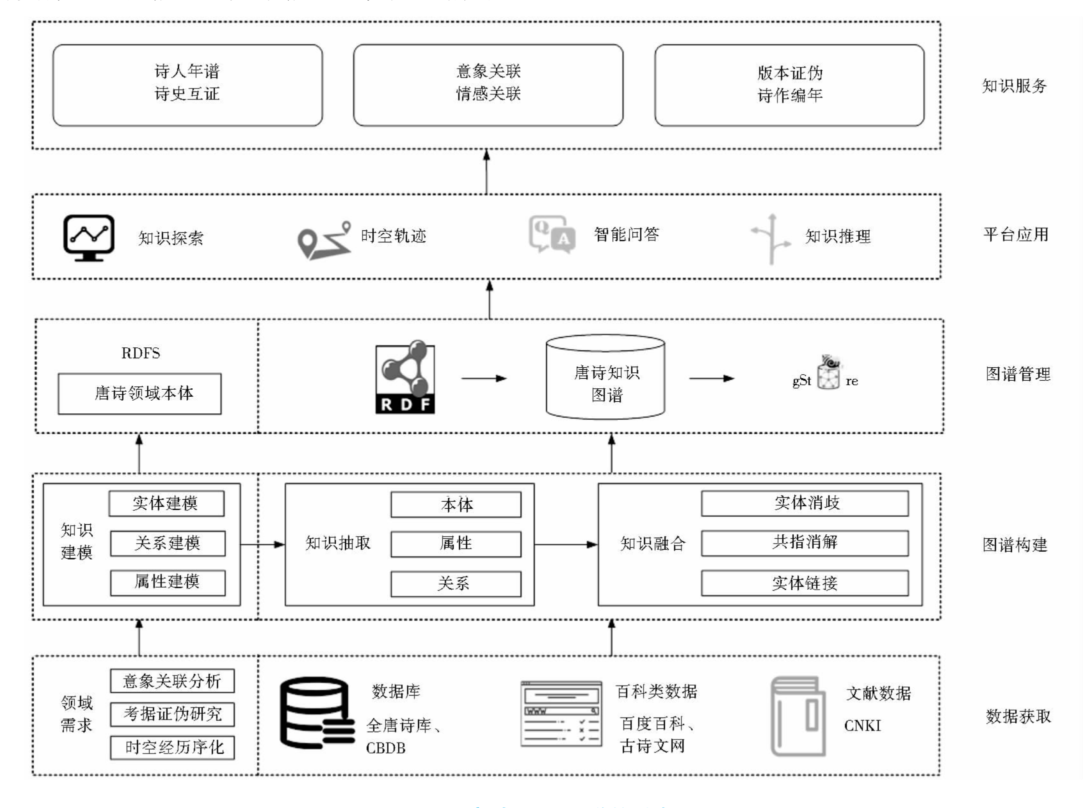


图 4　唐诗知识图谱构建框架

4　基于知识图谱的唐诗智能知识服务平台

基于第 3 部分构建的唐诗知识图谱，本文搭建智能化的唐诗知识服务平台 KnowPoetry（访问网址：http://kg.whu.edu.cn），提供对诗歌、诗人、经历等内在关联的全景式知识探索、时空轨迹、语义检索等应用，可辅助支持诗学、文献学和史学三大领域的唐诗研究，包括基于诗人时空轨迹可视化的史学研究、基于意象关联分析的诗学研究和基于诗歌版本证伪的文献学研究。

4.1　知识探索

基于唐诗知识图谱的知识探索能为用户提供更精准、更快速和更智能化的知识化内容资源，如诗人、诗

歌和经历等;能支持关联知识获取与分析服务,如诗人的引用关系探索、诗作的情感、意象、典故等维度的关联分析等。通过提供用户浏览检索的可视化界面,以环形图、力导图等形式为用户展示诗人、诗歌等多维知识图谱。用户可直接获取不同维度中的唐诗知识,并根据某些主题进行查询交互,从而实现知识导航的功能。

4.1.1 诗人主页 通过检索和定位某一位诗人可进入该诗人的主页,主页详细介绍诗人生平经历信息和创作的诗歌作品等,以力导图的形式展示了该诗人的作品图谱,并提供具体的节点信息。例如通过李白的作品图谱,我们可以了解其主要代表作品,并链接至作品主页,如图 5 所示:

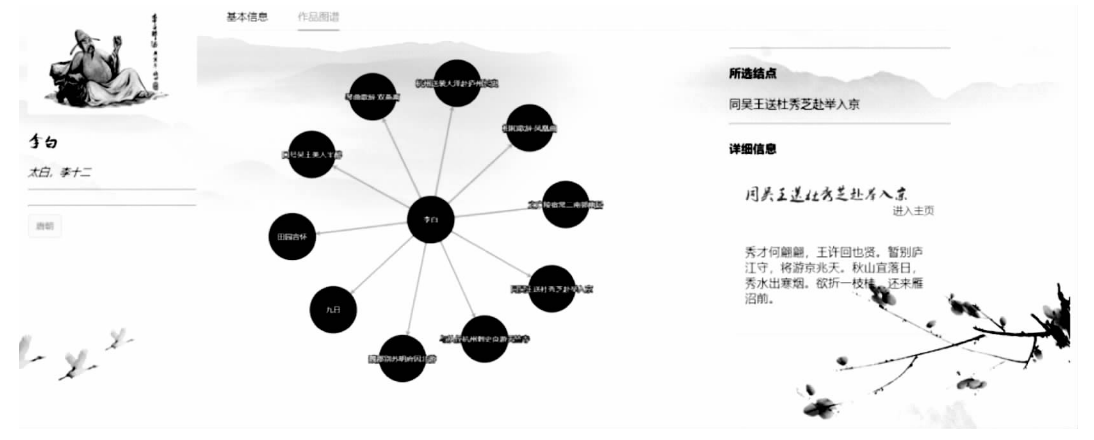


图 5 诗人主页示例:李白作品图谱

4.1.2 诗人引用关系图谱 通过定义诗人间的引用关系,本平台构建唐朝诗人社交网络,支持量化统计两位诗人间的互引次数,有利于了解诗人间的亲疏关系,衡量诗人的个人影响力。以诗人为主体,以其引用关系为链接形成了关联的诗人引用关系图谱,用户点击对应的节点可进一步了解诗人的详细信息,还可链接到其图片信息或诗人的个人主页。在图 6 中,可以看出与李白存在引用关系的诗人有:杜甫、高适、郑谷等,选择李白节点,则可以了解其详细信息。



图 6 诗人引用关系图谱

4.2 时空轨迹

历史上某位诗人所处的地理位置可能随着时间因素,因其升迁、贬黜、征战、归隐、游历、游学等经历而发生变化。因此从诗人对应的经历数据中,抽取人物、地点、时间和事件等属性类,将其轨迹动态映射到地图上。若诗人的某段经历数据残缺,我们可根据时空约束和诗人的创作经历计算最短路径,推测其轨迹。

如果仅是在地图中显示出诗人的全部轨迹,则其经历不免存在着诗人轨迹方向不明、起止地点不清晰等问题。遂将不同诗人的轨迹用不同的颜色进行标识,并将其全部轨迹进行了分段光标动态演示,以交互式地图形式帮助用户找寻在一定时期和某个地理区域

活动的诗人及其具体经历描述。如图 7 所示,该界面直观地揭示了李白由成都到扬州的一段时空经历及所作诗歌。

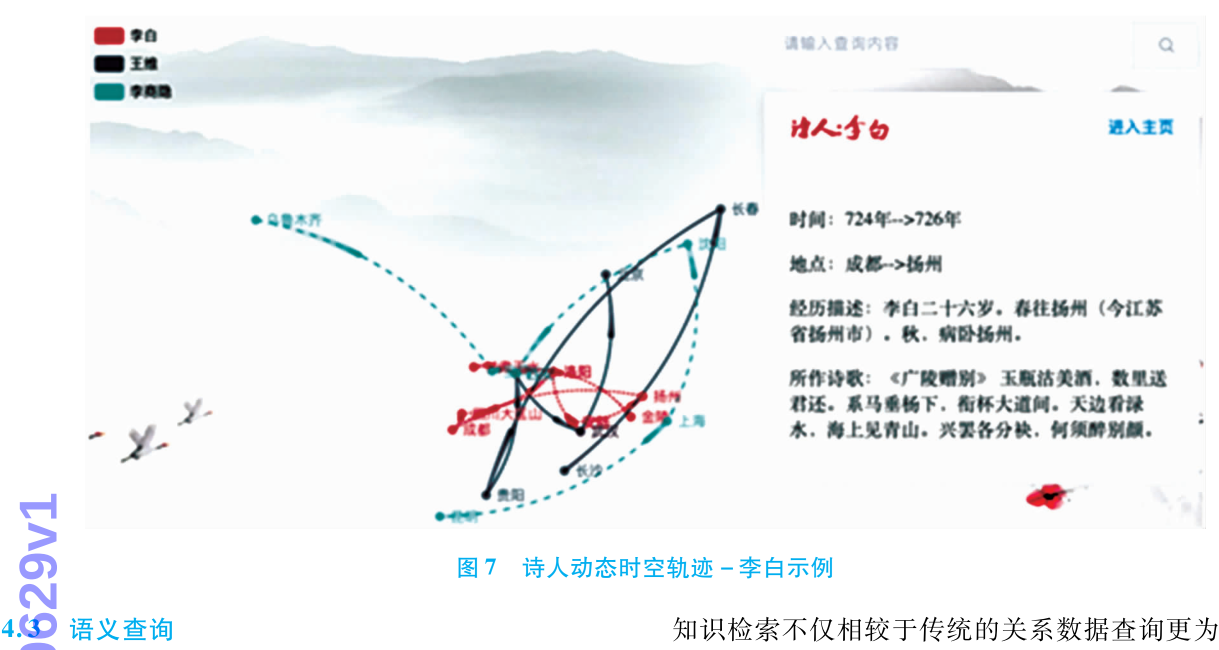


图 7　诗人动态时空轨迹－李白示例

4.3　语义查询

语义查询能够根据用户的自然语言问题,通过高效的子图匹配算法自动给出问题的答案。借助 RDF 数据库 gStore 进行查询操作,将 SPARQL 描述转化为子图匹配,使用高效的算法进行剪枝匹配,返回查询的结果。如处理"诗的标题含有白居易的诗"这一查询时,其 SPARQL 语言为:

```
sparql = ´
Select ? title ? name ? content
Where{
? poet <http://www.freekg.com/poet/name> ? name.
? poem <http: //www.freekg.com/poet/author> ? poet.
? poem <http: //www.freekg.com/poet/title> ? title.
Filter regex(? tags,\"^%s $ \")
? poem <http://www.freekg.com/poet/content>? content.}
% "白居易"
return sparql
```

在面对相对较为复杂的查询时,基于子图匹配的知识检索不仅相较于传统的关系数据查询更为快捷,也省去了需要用 SPARQL 查询语言调用多个库的繁琐操作,缩短了系统的响应时间[20]。如检索"白居易描写梅花的诗",其 SPARQL 语句为:

```
Sparql = ´
Select ? title ? name ? content
Where{
? poet <http://www.freekg.com/poet/name> ? name.
Filter regex(? name,"%s")
? poem <http: //www.freekg.com/poet/author> ? poet.
? poem <http: //www.freekg.com/poet/title> ? title.
? poem <http: //www.freekg.com/poet/tags> ? tags.
Filter regex(? tags,\"^%s $ \")
? poem<http://www.freekg.com/poet/content>? content.}
% ("白居易","描写梅花")
return sparql
```

系统会以较快的速度直接返回结果,并可以点击按钮进入诗歌主页。返回的 3 个结果如图 8 所示:

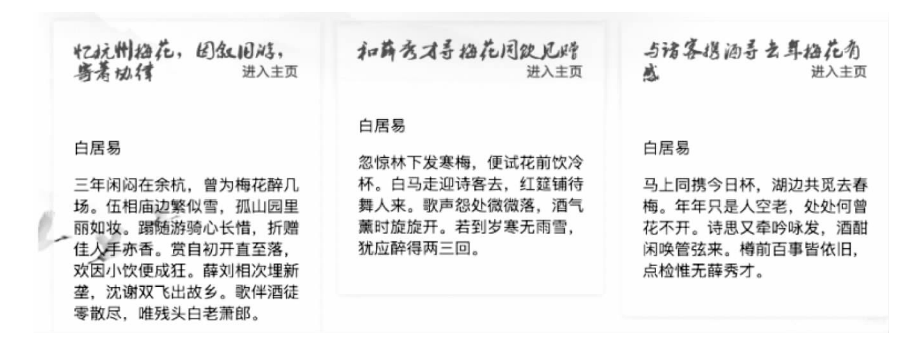


图 8　查询"白居易描写梅花的诗"结果

4.4 平台服务效果分析

通过构建唐诗知识图谱,我们可以回应大数据环境下唐诗领域的知识服务需求,搭建基于知识图谱的唐诗智能知识服务平台,实现数据驱动的唐诗领域知识研究的三大统一,形成一个知识创造的"智能生态循环系统",如图 9 所示。基于此循环系统,我们主要从研究方法、研究对象和研究目的三方面进行分析。

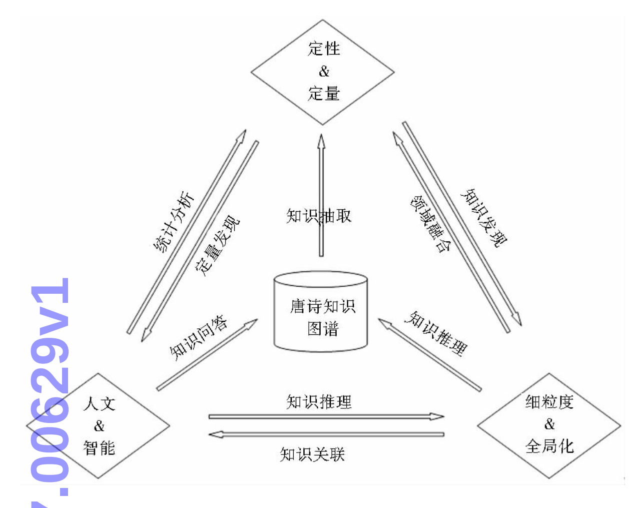


图 9 唐诗知识服务生态循环系统

(1)研究方法:定性与定量的统一。结合传统定性研究成果的基础,笔者通过构建知识图谱,可以量化地揭示唐诗中的诗人、诗歌、唐代历史纪年数据、空间地理坐标数据等知识,回答基于唐诗知识统计分析的研究问题,从而突破了唐诗研究的方法局限。例如回答李白最爱引用的典故是什么?通过对诗歌文本进行典故抽取与定量统计,得到李白诗作中引用典故次数高达 779 次,其中引用"巫山云"这一典故最多,有 11 次。

(2)研究对象:细粒度与全局化的统一。通过建立全局性的唐诗知识本体模型,我们可以将多源异构的唐诗数据有机地融合在一个相对完整全面的知识框架中,利用语义化的知识组织形式,便于机器读取和处理,实现基于诗歌、诗人和诗人经历的复杂关系的全局分析,从而扩宽了唐诗领域的研究对象。

(3)研究目的:人文与智能的统一。知识图谱利用自然语言处理技术,将复杂的自然语言问题转换为实体与实体间的关系,提供智能化的查询功能,为唐诗领域的数字人文研究提供知识服务,从而实现了唐诗数字人文研究范式创新的目的。

5 结语

以知识图谱为代表的人工智能技术,为唐诗研究和唐诗知识服务拓展了新方向。本文以"面向领域知识服务的唐诗知识图谱构建"为中心,遵循"需求分析→本体建模→知识库构建→知识服务平台设计"的研究思路,实现唐诗知识图谱的智能知识服务,包括知识探索、时空轨迹、语义查询、知识推理等。本文的研究一定程度上突破了传统数字人文单一化和细粒度的方法局限,为唐诗的研究与知识服务提供一种全新的、量化的、客观的、动态的视角。未来,我们希望能够通过扩充唐诗知识图谱,实现对诗歌和诗人数据更深层次的挖掘,推动人工智能环境下唐诗领域知识服务新形式的发展。

参考文献:

[1] DONG H. Enrichment of cross – lingual information on Chinese genealogical linked data[EB/OL]. [2018 – 12 – 02]. https://www.ideals.illinois.edu/handle/2142/98870.

[2] 徐增林, 盛泳潘, 贺丽荣, 等. 知识图谱技术综述[J]. 电子科技大学学报, 2016,45(4):589 – 606.

[3] 中国哲学书电子化计划[EB/OL]. [2018 – 10 – 15]. https://ctext.org/zh/.

[4] 中国历代人物传记数据库[EB/OL]. [2018 – 10 – 15]. https://projects.iq.harvard.edu/chinesecbdb.

[5] China Historical GIS[EB/OL]. [2018 – 10 – 15]. https://sites.fas.harvard.edu/ ~ chgis/.

[6] 搜韵:诗词门户网站[EB/OL]. [2018 – 10 – 15]. https://sou – yun.com/.

[7] MARKUS:中文古籍文本半自动标示平台[EB/OL]. [2018 – 10 – 15]. http://www.academia.edu/11078612/MARKUS_中文古籍文本半自動標記平台.

[8] 李友仁, 宋迎春. 北美与西欧的数字人文中国研究状况论析[J]. 山东社会科学, 2018(7):54 – 58,63.

[9] CHEN J W, BOROVSKY Z, KAWANO Y, et al. The Shishuo xinyu as data visualization[J]. Early Medieval China, 2014(20): 23 – 59.

[10] JASON P. Toward a Spatial History of Chan[J]. Review of religion and Chinese society,2016(3):164 – 188.

[11] 曾蕾, 王晓光, 范炜. 图档博领域的智慧数据及其在数字人文研究中的角色[J]. 中国图书馆学报, 2018, 44(1):17 – 34.

[12] 夏翠娟, 刘炜, 张磊, 等. 基于书目框架(BIBFRAME)的家谱本体设计[J]. 图书馆论坛, 2014,34(11):5 – 19.

[13] 夏翠娟. 中国历史地理数据在图书馆数字人文项目中的开放应用研究[J]. 中国图书馆学报, 2017,43(2):40 – 53.

[14] Desktop Poet[EB/OL]. [2018 – 10 – 15]. https://www.file – extensions.org/desktop – poet – file – extensions.

[15] 清华 AI 九歌计算机诗词创作系统[EB/OL]. [2018 – 10 – 15]. https://jiuge.thunlp.cn//.

[16] 猎户星在线写诗[EB/OL]. [2018 – 10 – 15]. http://www.dopoem.com.

[17] 微软亚洲研究院律诗绝句[EB/OL]. [2018－10－15]. http://duilian. msra. cn/jueju/.
[18] 微软六代小冰[EB/OL]. [2018－10－15]. http://www. msxiaobing. com.
[19] 唐诗别苑[EB/OL]. [2018－10－15]. http://poem. studentsystem. org.
[20] KVision 宋代学术语义网络[EB/OL]. [2018－10－15]. http://dh. kvlab. org/cbdb_kg/.
[21] 夏翠娟. 中国历史地理数据在图书馆数字人文项目中的开放应用研究[J]. 中国图书馆学报, 2017,43(2):40－53.

作者贡献说明：

周莉娜：负责论文主体内容撰写，包括引言、国内外相关研究、知识建模和唐诗智能知识服务平台等各部分内容的研究部分；
洪亮：指导论文整体框架思路，从宏观上把握唐诗知识图谱项目进展思路；
高子阳：负责论文中唐诗知识图谱构建部分内容的撰写，在唐诗知识图谱构建项目中负责数据管理和平台搭建的技术实现。

Construction of Knowledge Graph of Chinese Tang Poetry and Design of Intelligent Knowledge Services

Zhou Lina[1] Hong Liang[1] Gao Ziyang[2]
[1] School of Information Management, Wuhan University, Wuhan 430072
[2] College of Computer Science, Wuhan University, Wuhan 430072

Abstract: [**Purpose/significance**] Based on the demands of Tang poetry knowledge service under the current big data environment, the knowledge graph of Tang poetry is constructed and intelligent knowledge service is provided on the basis of large－scale data of Tang poetry, which promotes the innovation of knowledge management and knowledge service mode of Tang poetry under the artificial intelligence environment. [**Method/process**] Based on the investigation of domain knowledge service requirements, this paper designs the Tang poetry ontology model driven by domain knowledge service, then uses knowledge extraction, knowledge fusion, knowledge reasoning and other technologies to automatically construct the knowledge graph of Tang poetry which unifies the representation and organization of Tang poetry domain data and achieve semantic processing of large－scale Tang poetry data. [**Result/conclusion**] This paper designs an intelligent knowledge service platform KnowPoetry based on the Tang poetry knowledge graph, which provides intelligent knowledge services such as knowledge exploration, spatio－temporal trajectory, semantic query in the field of Tang poetry, and promotes the innovative transformation of Tang poetry digital humanities research methods in artificial intelligence environment.

Keywords: knowledge graph of Tang poetry　intelligent knowledge service　digital humanities　knowledge modeling